



## OPEN ACCESS

## EDITED BY

Kevin A. Henry,  
National Research Council Canada (NRC),  
Canada

## REVIEWED BY

San Hadzi,  
University of Ljubljana, Slovenia  
Patrick Chames,  
INSERM U1068 Centre de Recherche en  
Cancérologie de Marseille (CRCM), France

## \*CORRESPONDENCE

Charlotte M. Deane  
✉ deane@stats.ox.ac.uk

RECEIVED 30 May 2023

ACCEPTED 27 June 2023

PUBLISHED 18 July 2023

## CITATION

Gordon GL, Capel HL, Guloglu B,  
Richardson E, Stafford RL and Deane CM  
(2023) A comparison of the binding sites of  
antibodies and single-domain antibodies.  
*Front. Immunol.* 14:1231623.  
doi: 10.3389/fimmu.2023.1231623

## COPYRIGHT

© 2023 Gordon, Capel, Guloglu, Richardson,  
Stafford and Deane. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# A comparison of the binding sites of antibodies and single-domain antibodies

Gemma L. Gordon<sup>1</sup>, Henriette L. Capel<sup>1</sup>, Bora Guloglu<sup>1</sup>,  
Eve Richardson<sup>1</sup>, Ryan L. Stafford<sup>2</sup> and Charlotte M. Deane<sup>1\*</sup>

<sup>1</sup>Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford, United Kingdom, <sup>2</sup>Twist Bioscience, South San Francisco, CA, United States

Antibodies are the largest class of biotherapeutics. However, in recent years, single-domain antibodies have gained traction due to their smaller size and comparable binding affinity. Antibodies (Abs) and single-domain antibodies (sdAbs) differ in the structures of their binding sites: most significantly, single-domain antibodies lack a light chain and so have just three CDR loops. Given this inherent structural difference, it is important to understand whether Abs and sdAbs are distinguishable in how they engage a binding partner and thus, whether they are suited to different types of epitopes. In this study, we use non-redundant sequence and structural datasets to compare the paratopes, epitopes and antigen interactions of Abs and sdAbs. We demonstrate that even though sdAbs have smaller paratopes, they target epitopes of equal size to those targeted by Abs. To achieve this, the paratopes of sdAbs contribute more interactions per residue than the paratopes of Abs. Additionally, we find that conserved framework residues are of increased importance in the paratopes of sdAbs, suggesting that they include non-specific interactions to achieve comparable affinity. Furthermore, the epitopes of sdAbs are only marginally less accessible than those of Abs: we posit that this may be explained by differences in the orientation and compaction of sdAb and Ab CDR-H3 loops. Overall, our results have important implications for the engineering and humanization of sdAbs, as well as the selection of the best modality for targeting a particular epitope.

## KEYWORDS

single-domain antibody, antibodies, binding, paratope, epitope, structural biology, biologics

## 1 Introduction

Monoclonal antibodies are widely used as biotherapeutics, but their high molecular weight (~150 kDa) can cause high production costs as well as poor diffusion rates that limit tissue penetration (1–3). These properties of antibodies (Abs) have led to increased interest in recent years around smaller antibody fragments such as single-domain antibodies

(sdAbs). SdAbs are isolated VH domains (VHHs) homologous to the VH domain in antibodies and are derived primarily from camelid heavy-chain antibodies (4). SdAbs are approximately one tenth the mass of antibodies (~15 kDa). Given this smaller size, the structural diversity available to sdAbs is significantly reduced compared to Abs. However, sdAbs have been shown to achieve comparable binding specificities and affinities (5, 6). Furthermore, sdAbs are thermostable and have shown higher solubility, blood clearance and tissue penetration than Abs (2, 7, 8). These properties suggest that sdAbs have huge potential in therapeutic use, provided they can be successfully humanized (9).

Major structural differences exist between sdAbs and Abs, the most conspicuous being that sdAbs lack a light chain and therefore have only three complementarity-determining region (CDR) loops, half that of Abs. The CDR loops in both Abs and sdAbs are known to contain the majority of the binding site. Understanding the differences in the binding sites of these two classes of immunoglobulin, in terms of how their structures enable interaction with their binding partners, would facilitate decision-making as to which modality might be more effective when targeting a particular epitope.

In previous work, Zavrtnik et al. (2018) (6), suggested that sdAbs target more “rigid, concave, conserved and structured” epitopes. This hypothesis that sdAbs can target epitopes that are inaccessible to Abs is often linked to the fact that the CDR-H3 loops of sdAbs are longer than those of conventional Abs (10, 11). Zavrtnik et al. (2018) (6) and Mitchell and Colwell (2018a) (12) found an average difference in loop length of between three and four residues. Many papers have theorized that the longer CDR-H3 loops of sdAbs can protrude into concave spaces in a protein antigen surface that would be inaccessible to a conventional Ab with a shorter CDR-H3 loop (13–15). However, as highlighted by Henry and Mackenzie (2018) (16), isolated case studies make up much of the supporting literature on this idea. They note that “the degree to which sdAbs bind cryptic epitopes vs. conventional antibody-accessible epitopes ... remain[s] unknown.”

Aside from differences in CDR-H3 loop length, previous comparisons of the paratopes of sdAbs and Abs have shown that sdAbs have more hydrophobic character than Abs but are similarly enriched in aromatic residues (6). Furthermore, sdAbs tend to draw more residues from framework regions into the paratope, whereas Abs are more reliant on the CDR loops to interact with an antigen (Ag) (6, 12).

The previous studies of Zavrtnik et al. (6) and Mitchell and Colwell (12, 17) are limited by their relatively small datasets: Zavrtnik et al. analyze 105 sdAb-Ag crystal complexes, while Mitchell and Colwell compare sets of 90 sdAb-Ag and Ab-Ag crystal complexes (2018a) and then 156 sdAb-Ag and Ab-Ag complexes (2018b).

As sdAb datasets have increased in size in recent years (18), we have examined the binding sites of sdAbs and Abs using non-redundant datasets of 892 Ab-Ag and 345 sdAb-Ag structural complexes alongside non-redundant datasets of 1,614,526 human VH sequences [from Eliyahu et al., 2018 (19)] and 1,596,446 camel VHH sequences [from Li et al., 2016 (20)]. We find that in agreement with previous work, the paratopes of sdAbs are

smaller, on average, than those of Abs and that the CDR-H3 loop of sdAbs is longer. In our analysis, the paratopes of sdAbs and Abs show small differences in amino acid composition. We also find that the epitopes of sdAbs and Abs cannot easily be differentiated by their size, amino acid composition or accessibility. Overall, our results suggest that sdAbs and Abs do not target especially different epitopes, despite differences in their paratopes. However, they may be distinguishable by the manner in which they interact with these epitopes. We find that a greater number of interactions per residue are initiated by the CDR-H3 loop of sdAbs and that the framework region of sdAbs contributes more residues to the paratope. These differences likely contribute to the ability of sdAbs to achieve comparable binding affinity to Abs. However, our analysis shows that many of the binding framework residues are conserved positions, suggesting that sdAb binding may include non-specific interactions.

## 2 Methods

### 2.1 Sequence datasets

Non-redundant sequence datasets were obtained from the Observed Antibody Space (OAS) database (21). A set of 1,621,889 human VH sequences generated by Eliyahu et al. (2018) (19) and 1,601,636 camel VHH sequences generated by Li et al. (2016) (20), were filtered to remove duplicated sequences. Final datasets, referred to as the “Abs sequence dataset” and “sdAbs sequence dataset”, consist of 1,614,526 human VH sequences and 1,596,446 camel VHH sequences. These sequence datasets were used to compare the CDR lengths and the amino acid compositions of framework residues and CDR loops between Abs and sdAbs.

### 2.2 Structure datasets

We created up-to-date, non-redundant datasets of both Abs and sdAbs that were in complex with protein antigens (Ags). We refer to these as the “Abs structural dataset” and “sdAbs structural dataset”. These structures were extracted from SABDab (22) and SABDab-nano (18) on the 23<sup>rd</sup> February 2022. The datasets were extracted as follows:

1. Only Ab-Ag and sdAb-Ag complexes for which at least one of the CDR residues of the antibody is in close contact, defined as under 4.5 Å, to the antigen.
2. Only the Abs and sdAbs identified as in a complex with a protein antigen (< 50 residues), according to SABDab annotations.
3. Only structures of complexes solved by X-ray crystallography to  $\leq 3.0$  Å resolution.
4. Abs and sdAbs were filtered separately to remove redundancy using a sequence identity cut-off of 95% across the IMGT-defined CDR residues using CD-HIT (23).
5. A small number of complexes were reintroduced if their epitope identity score was less than 75% compared to any

other complex, to include complexes containing similar CDRs but different epitopes. To calculate epitope identity, epitope sequences were first aligned using CD-HIT. Based on the aligned positions, the epitope identity score was determined as the fraction of matching (distance-defined) epitope residues (same amino acids and same aligned position) across the epitope residues of the two antigens.

The resulting sdAbs structural dataset consisted of 345 complexes, of which 309 had “unique” CDRs. The final Abs structural dataset consisted of 892 complexes, of which 792 had “unique” CDRs. [Supplementary Text S1](#) and [Table S1](#) give further detail on dataset curation and a breakdown of the number of complexes remaining at each filtering step. [Table S2](#) shows species variation for both structural datasets. [Supplementary Figure S1](#) shows distributions of epitope identity across datasets.

## 2.3 Numbering definitions

The IMGT numbering scheme and CDR definitions were used throughout this work (CDR1: IMGT residues 27-38, CDR2: IMGT residues 56-65, CDR3: IMGT residues 105-117) ([24](#)). ANARCI ([25](#)) was used to number all of the Abs and sdAbs.

## 2.4 Binding site definitions

We describe the binding site using three definitions. As used in most methods annotating and predicting paratopes or epitopes, we consider a distance definition, which includes all antibody residues which are in close contact with the antigen ( $\leq 4.5$  Å). A very similar result is achieved by defining the binding site by solvent-accessible surface area (SASA), where residues are included in the paratope or epitope if they become buried on complex formation (SASA-defined). In our work we focus on defining the binding site by the interactions occurring between pairs of residues, using Arpeggio ([26](#)). Arpeggio determines interaction types based on distance, angle, and atom type. It was run on each PDB file in both structure datasets after cleaning with the associated cleaning script<sup>1</sup>, using a distance threshold of 4.5 Å. This generates a five-bit fingerprint for each pairwise interatomic contact which shows the type of interactions occurring. These include, van der Waals, steric clashes, covalent bonds, proximal interactions (defined as being within the cut-off distance but not representing a meaningful interaction) and specific interactions such as hydrogen bonds. This output was processed to exclude interactions with water molecules and chains other than the antibody and antigen. Heterogens were removed with BioPython ([27](#)). For all remaining positions, the interatomic interactions were summarized per residue-residue pair. Residues were considered to interact if at least one of the atom-atom pairs in these residues established a van der Waals (vdW) bond or a specific interaction. Clashing vdW and proximal interactions were

classified as contacts if no specific bonds were observed. We refer to this latter definition of the binding site as the interactions-defined paratope and interactions-defined epitope.

Interatomic interactions between the Ab-Ag and sdAb-Ag complexes were compared by counting the total number observed. If multiple interaction types were identified between a single pair of atoms, the interactions were counted individually. Mean and standard deviation of the observed interactions were calculated by sub-sampling 10% of the total set of interactions 1000 times.

[Supplementary Figures S2, S3](#) visualize the difference between paratopes and epitopes defined by contacts or interactions, and the difference between each definition of the binding site.

## 2.5 Amino acid composition

The sequence datasets were used to compare compositions of CDR loops. The sdAbs and Abs sequence datasets were split by germline and only those belonging to IGHV3 compared: this included all sequences for the sdAbs dataset but reduced the Abs dataset to 761,235 sequences. Sequences were aligned using ANARCI numbering annotation. The proportions of individual amino acids at each position in each CDR-H loop were determined. Positions were omitted where less than 5% of sequences had an amino acid at that position.

To assess the conservation of framework residues that appear in the paratope, firstly the structural datasets were used to determine which positions are often involved in the paratopes of sdAbs and Abs. Framework residues were considered as important contributors to the paratope if they were observed in at least 10% of the complexes in our datasets. The amino acid compositions of these same positions were then obtained from the sequence datasets as a background for comparison.

## 2.6 Epitope accessibility

Multiple methods are available that describe the curvature of a surface. However, these methods struggle to successfully capture the complex nature of the epitope surface. Here, we have designed a simple metric using the solvent accessible surface area to compare the accessibility of the epitopes targeted by sdAbs and Abs.

We define “epitope accessibility” as the solvent accessible surface area (SASA) of the epitope surface relative to the sum of the SASA values of the epitope residues as if they were isolated in space. The function “get\_sasa\_relative” from the PyMOL cmd package ([28](#)) was used to calculate the SASA values, where residues with a value of 0 are considered completely buried, and those with a value of 1 are completely exposed. As such, the sum of the SASA of epitope residues were they to be isolated in space is equivalent to the total number of residues in the epitope: this is reflected in our implementation of the metric. Differences in the distributions of epitope accessibility for sdAbs and Abs are determined via bootstrap re-sampling.

<sup>1</sup> <https://github.com/harryjubb/pdbtools>

## 2.7 Canonical forms of the CDRs

Canonical forms of sdAb and Ab structures were identified using the PyIgClassify2 database (29).

## 2.8 Structural clustering

Antibody chains from the 345 sdAb-Ag and 892 Ab-Ag complexes were extracted, giving 301 and 838 unique sdAbs and Abs structures (as some PDB entries include sdAbs or Abs that form complexes with multiple antigens). A greedy clustering method was used where each of the sets of CDR-H1, CDR-H2 and CDR-H3 loops were clustered based on their length and RMSD with a cut-off of 1.5 Å. The number of clusters which contain both sdAbs and Abs (overlap clusters) was determined. The expected number of overlap clusters was found by generating random clusters of matching size. Random clusters were generated 20 times from the original set of all Ab and sdAb structures and the mean and standard deviations for the number of overlap clusters was calculated.

## 2.9 Orientation of CDR-H3 loops

We analyzed the general orientation of the CDR-H3 loops of Abs and sdAbs by examining their centers of geometry in reference to an  $R^3$  coordinate system (see [Text S2](#) for method and [Supplementary Figure S4](#)). The dataset used for this analysis includes the structures of 388 bound sdAbs, 116 unbound sdAbs, 1977 bound Abs and 862 unbound Abs. Structures were downloaded from SAbDab (22) on 8<sup>th</sup> August 2022 and generated individually to be non-redundant at 95% sequence identity. Structures were numbered with the IMGT scheme using ANARCI (25) and CDR definitions used accordingly. Any structures with missing backbone atoms in CDR-H loops or anchors (three residues on either side of each loop) were also removed.

Using the spherical coordinates method,  $\rho$  describes the reach of the CDR-H3 loop away from the rest of the VH domain. A CDR-H3 loop in an extended conformation will have a high  $\rho$  value whereas a loop of identical length that is folded against the VH domain will have a lower value.  $\phi$  gives an indication of whether the CDR-H3 loop is horizontally oriented towards the rest of the VH domain or away from it. In the case of Ab structures, a high  $\phi$  value indicates packing against the VL domain.  $\theta$  gives a measure of the elevation of the loop. A low value corresponds to a CDR-H3 that extends directly up and away from the rest of the VH domain, whereas a high value indicates that the loop is “folding” down. In the case of Ab structures, a high  $\theta$  value corresponds to a loop that is packed into the groove created by the VH-VL interface. Lastly, we divide the loop length by  $\rho$  to give a measure of compaction. A loop with low compactness uses its entire length to reach away from the VH domain, whereas high compactness corresponds to a loop that is packed against the VH.

## 2.10 Statistical tests

As not all distributions followed the normal distribution, significant differences between the sdAbs and Abs were tested by bootstrap re-sampling in which 5000 bootstrap samples are taken of size 300. The unpaired mean difference and the p-value of the two-sided permutation t-test are reported. Results are described as significant for p-value < 0.05.

## 2.11 Visualizations

All visualizations were created using open-source PyMOL v2.4.1 (28), UCSF ChimeraX (30), or matplotlib v3.5.1 (31).

## 3 Results

In this study, non-redundant sequence datasets for Abs and sdAbs of size 1,614,526 and 1,596,446 respectively, and non-redundant structural datasets of 892 Ab-Ag and 345 sdAb-Ag complexes, were compared with respect to their paratopes, epitopes and their interactions with their respective antigens to identify the differences and similarities between their binding sites, and to determine whether these two modalities target different types of epitopes.

### 3.1 The CDR-H3 loop is longer in sdAbs than in Abs

Previous work has shown that the CDR-H3 loops of sdAbs are longer than those of Abs. Lengths of the CDR loops were compared for both sequence and structural datasets. When comparing the sdAbs and Abs sequence datasets, we find that the CDR-H1 loops of Abs are, on average, slightly longer than those of sdAbs by 0.4 residues. Abs have on average longer CDR-H2 loops by 0.2 residues. The CDR-H3 loops are significantly longer in sdAbs by 1.4 residues on average ([Figure 1A](#)). The results from the structural dataset are consistent with the trends observed for the sequence datasets: for the solved structures, bootstrap re-sampling shows that for CDR-H1, there is a significant difference between sdAbs and Abs of 0.2 residues. For CDR-H2, we find that there is a difference of 0.08, however this was not significant (p-value = 0.12). For the structural datasets, the CDR-H3 loop is significantly longer in sdAbs than in Abs by 1.6 residues ([Figure 1B](#)). This finding agrees with previous studies.

### 3.2 Structural clustering shows a separation between Abs and sdAbs CDR structures

Further to comparing the lengths of the CDR loops found in Abs and sdAbs, we next structurally clustered the CDR loops to determine whether they adopt distinct conformations and occupy

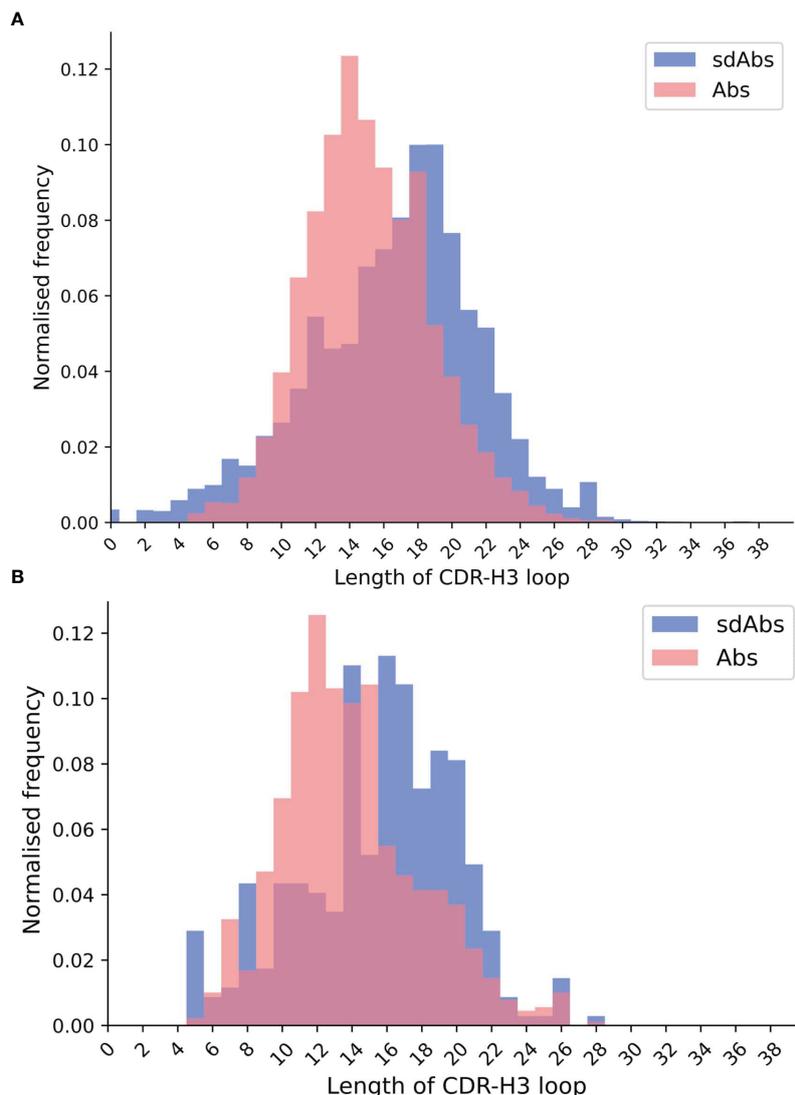


FIGURE 1

The distributions of CDR-H3 loop length for (A) sequence data and (B) structural data both show that CDR-H3 loops in sdAbs (blue) tend to be longer than those in Abs (pink).

different regions of structural space. If Abs and sdAbs were to adopt different paratope shapes, this would suggest that the epitopes they are able to bind would differ.

Our initial approach was to assign canonical forms to each of the Abs and sdAbs loop structures, according to updated canonical forms from Kelow et al. (2022) (29). However, for both Abs and sdAbs a significant percentage of CDR loops could not be assigned a canonical form. Therefore, CDR loops were clustered based on length and RMSD, with a cut-off of 1.5 Å. Clustering of the CDR loops of our 838 Abs and 301 sdAbs structures collectively returned 168 clusters for CDR-H1, 94 clusters for CDR-H2 and, as expected given the differences in CDR-H3 length and the high variability of CDR-H3 in general, 729 CDR-H3 clusters.

The number of clusters containing both Abs and sdAbs structures was determined and a mean and standard deviation for the expected number of overlap clusters, if random clustering had occurred, was calculated (Table 1). For CDR-H1, 18 clusters

contained both Abs and sdAbs compared to an expected value of  $16.2 \pm 1.29$  for random clusters. For CDR-H2, 23 clusters contained both Abs and sdAbs compared to an expected value of  $22.3 \pm 0.829$ . For CDR-H3, there were 10 overlap clusters compared to an expected value of  $3.30 \pm 1.55$ . Overall, we observe that for the CDR-H1 and CDR-H2 loops, the number of clusters we see with both Abs and sdAbs occurring within them is within the range of what would be expected had the structures been clustered at random. This indicates that sdAbs and Abs may assume distinct CDR conformations. As the CDR loops form the majority of the binding site, this suggests that Abs and sdAbs may prefer to bind in different ways.

### 3.3 SdAbs and Abs have more identical CDR sequences than expected by chance

We next examined the CDR loop sequences belonging to IGHV3 germlines, taken from the sdAbs and Abs sequence

**TABLE 1** Clustering the structures of the CDR loops of sdAbs and Abs based on length and conformational similarity (measured by RMSD) shows that there is overlap between the shapes that CDR loops of sdAbs and Abs form. However, this overlap is within the range of that observed on random clustering, and as such suggests that sdAbs and Abs adopt distinct CDR conformations. Values in the table show the number of structures within each cluster, with the number of clusters containing only a single structure shown in brackets.

	CDR loop		
	CDR-H1	CDR-H2	CDR-H3
<b>Abs-only</b>	66 (48)	35 (21)	489 (383)
(single-occupancy)			
<b>SdAbs-only</b>	84 (64)	36 (25)	230 (197)
(single-occupancy)			
<b>Overlap</b>	18	23	10
<b>Total</b>	168	94	729

datasets. This reduced the size of the Abs dataset to 761,235 sequences (all 1,596,446 sequences in the sdAbs sequence dataset belong to the IGHV3 germline). Sequences within each dataset were aligned via ANARCI annotation and the amino acid composition at each position in each loop determined. Positions were omitted if less than 5% of sequences in a dataset had a residue at that position. [Supplementary Figure S5](#) shows sequence logo plots of the CDR loops of Abs and sdAbs.

Given the size of the sequence space, the probability of finding the same sequences in both Abs and sdAbs CDR loops is low. The expected proportion of identical sequences between the sdAbs and Abs sequences for each loop was calculated and compared to the actual overlap. For CDR-H1, the expected overlap is  $6.31 \times 10^{-11}$  versus 0.024, for CDR-H2,  $7.33 \times 10^{-11}$  versus 0.021, and for CDR-H3,  $1.53 \times 10^{-21}$  versus  $3.00 \times 10^{-4}$ . As the actual number of identical sequences is greater than the expected number, this suggests that there are similarities in the amino acid compositions of sdAbs and Abs CDR loops, which likely arise from their similar genetic background.

### 3.4 Paratopes of sdAbs and Abs show small differences in their amino acid compositions

In addition to assessing differences in the CDR loops of Abs and sdAbs, we considered whether there are overall differences in their respective paratopes by firstly comparing their amino acid composition. Following the work of Wong et al. (32), amino acid compositions for the paratopes were determined by classifying amino acids into seven classes (aliphatic, aromatic, sulfur, hydroxyl, basic, acidic and amine). For each paratope, the fraction of each observed class was determined and the distributions of amino acid types for paratopes of sdAbs and Abs were compared.

Comparisons of the seven classes reveal that, for both distance-defined and interactions-defined paratopes, there are small increases in the proportions of aliphatic, sulfur and basic residues in sdAb paratopes ([Supplementary Figures S6, S7](#)). We observe a decrease in aromatic residues in sdAb paratopes. There are no significant differences in the proportions of residues in the hydroxyl, acidic or amine classes.

### 3.5 SdAbs paratopes are significantly smaller than those of Abs

Next, we compared the sizes of sdAbs and Abs paratopes. Here, we define size by the number of residues in the paratope. Previous work has revealed that sdAbs can show comparable binding affinity to Abs despite their smaller size (5, 6). Given that sdAbs are missing the VL domain and therefore half of an Ab potential binding site, we would expect them to also have a smaller paratope. Using our non-redundant structural datasets, we compared the size of sdAb and Ab paratopes for each of the distance-defined, interactions-defined and SASA-defined paratopes. We found that for distance-defined paratopes, sdAb paratopes are significantly smaller than Ab paratopes by 3.6 residues and for interaction-defined paratopes, SdAb paratopes are smaller than Ab paratopes by 2.6 residues ([Figure 2](#)). [Supplementary Figure S8](#) shows results consistent with the above for the SASA-defined paratopes. The differences found between the CDRs and more specifically the paratopes of sdAbs and Abs in our datasets suggest that these two modalities may target distinct epitopes.

### 3.6 Epitopes targeted by sdAbs and Abs have similar amino acid compositions

We next assessed the epitopes of Abs and sdAbs. One factor that may differ between sdAbs and Abs is the amino acid compositions of their target epitopes. As for the paratope amino acid compositions, amino acid compositions for the epitopes were determined by classifying amino acids into seven classes (aliphatic, aromatic, sulfur, hydroxyl, basic, acidic and amine).

Comparisons of the seven classes for both distance-defined and interactions-defined epitopes show that for epitopes of sdAbs, there is a small but significant increase in the number of aromatic residues, and a significant decrease in the number of basic residues ([Supplementary Figures S9, S10](#)). Given that Abs and sdAbs are a highly similar class of molecules, it would be expected that differences in the epitope amino acid compositions would be minimal. Our results reflect this: significant differences are found but these are minor in the absolute sense. Thus, we conclude

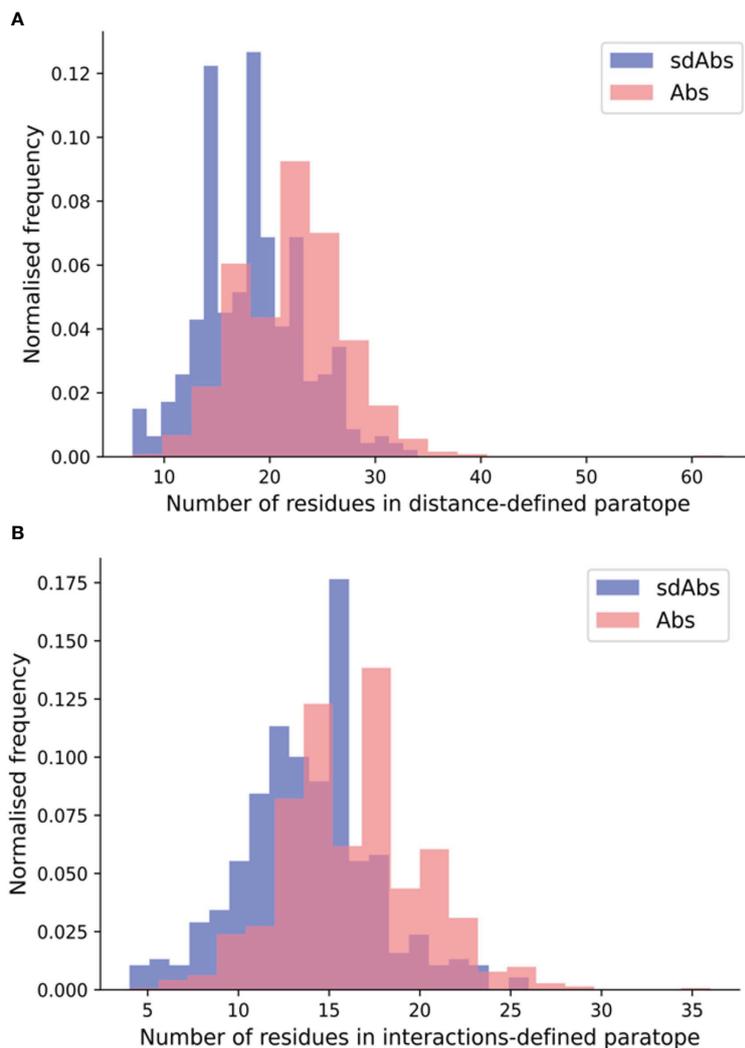


FIGURE 2

The paratopes of sdAbs (blue) tend to contain fewer residues than the paratopes of Abs (pink). (A) Distributions of the number of residues in the distance-defined paratopes, where sdAbs paratopes contain significantly fewer by 3.6 residues on average compared to Abs. (B) Distributions of the number of residues in the interactions-defined paratopes, where sdAbs paratopes contain significantly fewer by 2.6 residues, on average.

that the epitopes of sdAbs and Abs are difficult to distinguish between based on their amino acid composition.

### 3.7 Epitopes of Abs are more linear than those of sdAbs

Epitopes are often characterized by whether they are more linear or discontinuous in nature. A linear epitope is formed from amino acid residues that fall next to each other at the primary sequence level, whereas a discontinuous epitope is formed from residues that are not adjacent in the amino acid sequence but are pulled together upon folding (33, 34). Here, we determined whether Abs and sdAbs show distinct epitope preferences in terms of epitope continuity. We represent how continuous an epitope is by the number of contiguous residues in the epitope sequence.

For both the distance and interactions-based definitions, epitopes of Abs tend to be slightly more linear than those of

sdAbs (Figure 3). Abs showed a significantly greater percentage of linear residues for both the distance-defined (4.6%) and interactions-defined (6.9%) epitopes. Similar results are observed when comparing the raw count of linear residues (Supplementary Figure S11). Results are replicated for the SASA-defined epitopes (Supplementary Figure S12). As the epitopes of sdAbs and Abs are of comparable size, the fact that Abs have slightly more linear epitopes than sdAbs is not due to a difference in epitope size.

### 3.8 Epitopes targeted by sdAbs and Abs are of comparable size

When size is defined by the number of residues, the paratopes of sdAbs are smaller than those of Abs, which suggests that sdAbs may be limited to binding smaller epitopes. Here, we determined the

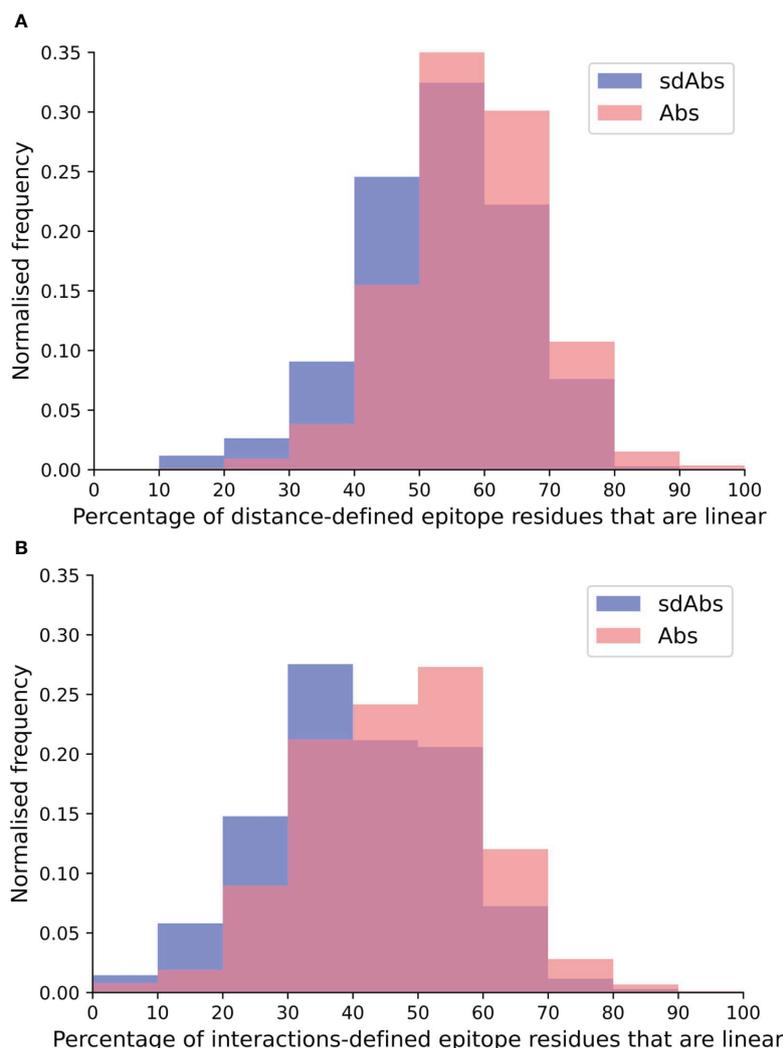


FIGURE 3

The epitopes targeted by Abs are relatively more linear than epitopes targeted by sdAbs, as suggested by the distributions of percentages of linear residues for epitopes targeted by Abs (pink) and sdAbs (blue) for the (A) distance-defined epitopes and (B) interactions-defined epitopes.

number of residues in the distance-defined epitopes, the SASA-defined epitopes and the interactions-defined epitopes for our non-redundant structural datasets. Our results show that for each of our epitope definitions, there is no significant difference between the size of the epitopes targeted by sdAbs and Abs (Figure 4, Supplementary Figure S13). Despite their smaller paratope size, sdAbs target epitopes of equal size to those targeted by Abs. This indicates that the paratopes of sdAbs must interact with their epitopes in a different way to that of Abs paratopes.

### 3.9 Epitopes targeted by sdAbs and Abs are of similar accessibility

In agreement with existing studies on smaller datasets, we found that sdAbs have longer CDR-H3 loops than Abs. Previous work has suggested that this facilitates interactions between sdAbs and

epitopes that are less accessible to conventional Abs (5, 11, 13–15). To assess whether the epitopes of sdAbs do indeed tend to be less accessible, the accessibility of all interaction-defined epitopes of sdAbs and Abs was analyzed.

We define epitope accessibility as the total relative SASA for the epitope surface, divided by the sum of the relative SASA values for each epitope residue were they completely exposed (equivalent to the number of residues in the epitope).

We found that the epitope accessibility of sdAbs was significantly lower than that of Abs: the unpaired mean difference between the epitope accessibility of sdAbs and Abs was 0.046 (Figure 5). These results support previous studies that suggest that sdAbs are able to target epitopes that are inaccessible to Abs (6). There is however also a large overlap in the distributions, and the absolute difference is small: this supports the suggestion from Henry and MacKenzie (2018) (16) that there is likely overlap in the types of epitopes that sdAbs and Abs target.

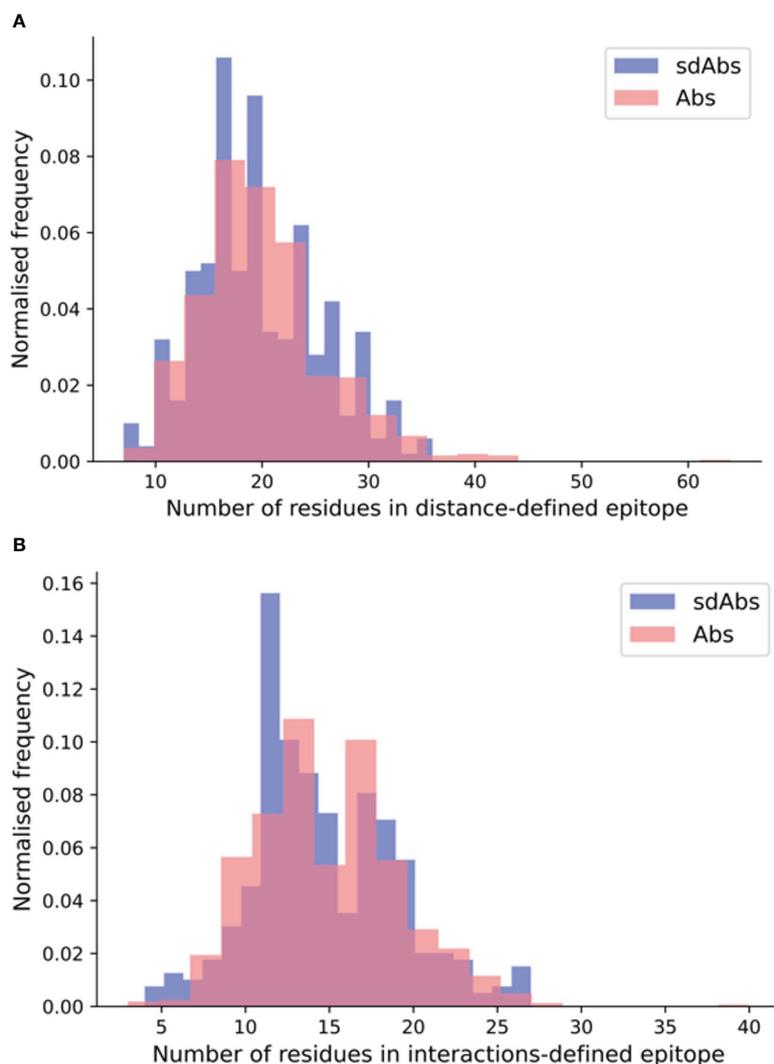


FIGURE 4

SdAbs are able to target epitopes of equal size (as defined by number of residues) to those targeted by conventional Abs, as suggested by the distributions of the number of residues in the (A) distance-defined epitopes for Abs (pink) and sdAbs (blue) structural datasets, where a mean difference of 0.59 is observed between sdAbs and Abs ( $p$ -value = 0.22) and (B) interaction-defined epitopes, where a mean difference of 0.32 is observed between sdAbs and Abs ( $p$ -value = 0.34).

### 3.10 CDR-H3 loop length does not correlate with epitope accessibility

The hypothesis that sdAbs are generally able to target epitopes that are less accessible to conventional Abs derives from the finding that their CDR-H3 loops are longer than those of Abs (10, 11). However, there is no correlation between the length of the CDR-H3 loop and the epitope accessibility for our datasets (Figure 6). For sdAbs, the Pearson correlation coefficient for epitope accessibility against the CDR-H3 loop length was -0.021. For Abs, the Pearson correlation coefficient for epitope accessibility against the CDR-H3 loop length was -0.097. These results indicate that the length of the CDR-H3 loop alone does not influence the accessibility of the epitope targeted by either antibody type.

### 3.11 Abs and sdAbs target epitopes of similar accessibility due to packing of sdAb CDR-H3 loops against the VHH domain

In light of our finding that the length of the CDR-H3 loop does not dictate the accessibility of the epitope to which a paratope binds, we examined the differences in the orientation of Ab and sdAb CDR-H3 loops relative to the rest of the VH domain, to determine how the conformation of the CDR-H3 loop may affect epitope preference.

We use four descriptors to describe the orientation of the CDR-H3 loops (see Methods, Supplementary Text S2 and Supplementary Figure S4): the parameter  $\rho$  represents the reach of the CDR-H3 loop away from the VH domain,  $\phi$  describes the horizontal

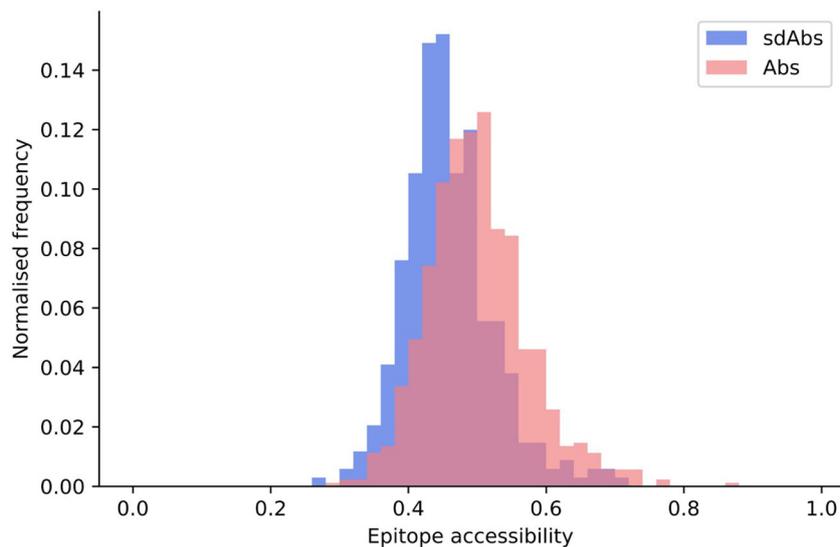


FIGURE 5

Epitopes targeted by sdAbs are slightly less accessible than those targeted by Abs. Distributions of epitope accessibility for the interactions-defined epitopes of sdAbs (blue) and Abs (pink) were found to be significantly different, though the absolute difference is small: the unpaired mean difference between sdAbs and Abs epitope accessibility was 0.046.

orientation of the CDR-H3 towards the rest of the VHH (for sdAbs), or against the VL domain (for Abs),  $\theta$  describes loop extension where a low value corresponds to a CDR-H3 extending up and away from the rest of the VH domain and lastly compaction, which is determined by dividing loop length by  $\rho$ .

Near-identical distributions of  $\rho$  values suggests that the two types of antibodies have similar reach, indicating that sdAbs cannot necessarily provide extended paratopes via their CDR-H3 loops compared to Abs (Figure 7A). A shoulder in the distribution of  $\rho$  values for Abs above the median value suggests that Abs may be more able to target deeper epitopes that require a longer reach.

The observation that sdAb CDR-H3 loops tend to be longer than those in Abs, whilst having similar reach, may be explained by loop compaction. On average, sdAb CDR-H3 loops are much more compacted than Ab loops (Figure 7B). The distribution of compactness scores for sdAbs is bimodal, with the first peak corresponding to the distribution found in Abs. This suggests one population of sdAb CDR-H3 loops that behaves similarly to Ab CDR-H3 loops, and one population that is more folded against the VHH domain (Figure 8A). SdAbs can either increase their reach with CDR-H3 length at a rate similar to Abs, or their loops can remain in a more heavily compacted state.

Compared to Ab CDR-H3 structures, sdAbs show a much wider bimodal distribution of  $\theta$  values, with the major peak corresponding to  $\theta$  values in excess of those observed for Ab structures, and another minor peak below the Ab distribution (Figure 7C). This indicates that the majority of sdAb CDR-H3 loops lie flat against the rest of the VHH domain, therefore folding down. We observe a slight shift in  $\theta$  in the distribution for bound sdAbs, but note that the position of the peaks still remains stable. We conclude that sdAbs generally do not extend their CDR-H3 loops upon binding, as has previously been hypothesized. Lastly, we

find near-identical values of  $\phi$  for sdAbs and Abs, with sdAb  $\phi$  values having a slightly wider distribution (Figure 7D).

To examine how CDR-H3 loops pack against the VH or VL domains, we analyzed the relationship between the spherical angles and compactness. Both sdAb and Ab CDR-H3 loops become more compacted through an increase in  $\theta$ : packing of the loop down towards the rest of the VH domain decreases its reach (Figure 8B). We hypothesize that this is a mechanism to stabilize the paratope structure by allowing the loop to pack against the rest of the VH domain. We also find an inverse relationship between compactness and  $\phi$  for sdAbs and Abs (Figure 8C). As  $\phi$  increases (as the CDR-H3 loop is horizontally oriented away from the VH domain), sdAbs show an increase in compactness whereas the opposite is true for Abs. For sdAbs, an increase in  $\phi$  results in the loop extending away into empty space, whereas in Abs the loop is positioned towards the VL domain. As the presence of the VL domain provides steric hindrance, the CDR-H3 loop is forced into a conformation that orients it away from the Ab, therefore reducing compactness and increasing reach.

### 3.12 SdAbs establish more interactions with their epitope per paratope residue than Abs

Our results thus far demonstrate that there are differences between the paratopes of sdAbs and Abs. But, our results also find only limited differences between the epitopes of the two modalities.

We have shown that for our datasets, Abs and sdAbs are able to bind similarly-sized epitopes, despite sdAbs paratopes being smaller. In order to investigate how this is achieved, we compare the

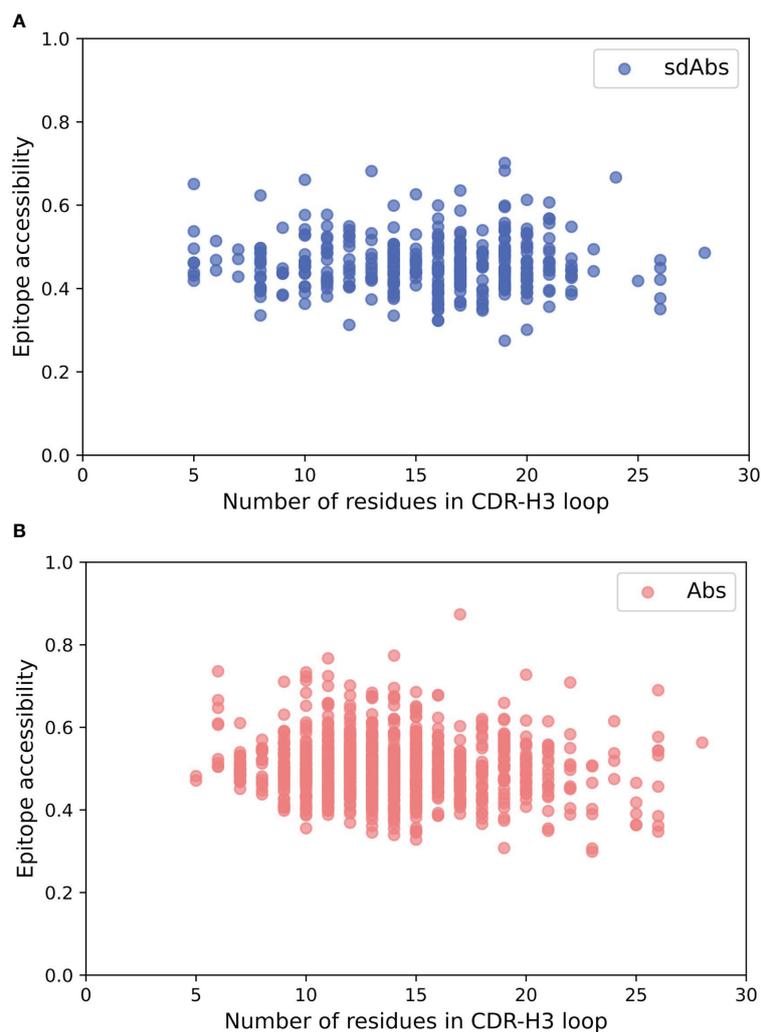


FIGURE 6

There is no correlation between the length of the CDR-H3 loop and the accessibility of the epitope surface for either Abs or sdAbs. (A) Correlation between accessibility of sdAb epitopes and length of CDR-H3 loop. (B) Correlation between accessibility of Ab epitopes and length of CDR-H3 loop.

interactions observed within binding sites. We find that, normalizing for the size of the paratope, per paratope residue, sdAbs establish significantly more interactions than Abs (Figure 9). This suggests that sdAbs establish a similar binding affinity to Abs by each paratope residue having an increased number of interactions with the epitope.

### 3.13 Hydrophobic interactions dominate both sdAb-Ag and Ab-Ag complexes

As well as the number of interactions, the types of interactions established between the antigen and the antibody in sdAbs and Abs were compared. All interatomic interactions between the interaction-defined epitope and paratope residues were considered. Each type of interaction was counted individually if an atom-atom complex established more than one interaction type (see Methods for full details).

In terms of interactions arising from the CDR loops, very similar types are observed (Figure 10A), whilst for the framework

regions involved in binding, we see an increase in hydrophobic interactions for sdAbs compared to Abs and the VH domain of Abs alone (Figure 10B).

### 3.14 CDR-H3 and framework residues are of increased importance for interactions in the sdAb-Ag complex

Next, we compared the relative contributions of the CDR loops to interactions within the binding site, including the mean number of interactions per loop (Supplementary Figure S14). In our data, we see the expected dominance of the CDR-H3 loop in binding. We found that there are significantly more interactions contributed from the CDR-H3 in sdAbs than Abs (Supplementary Figure S15A) even after normalizing for CDR-H3 length (Supplementary Figure S15B) and that in sdAbs, there was a significantly greater contribution from the CDR-H3 residues both in terms of contributing residues to the paratope and contributing

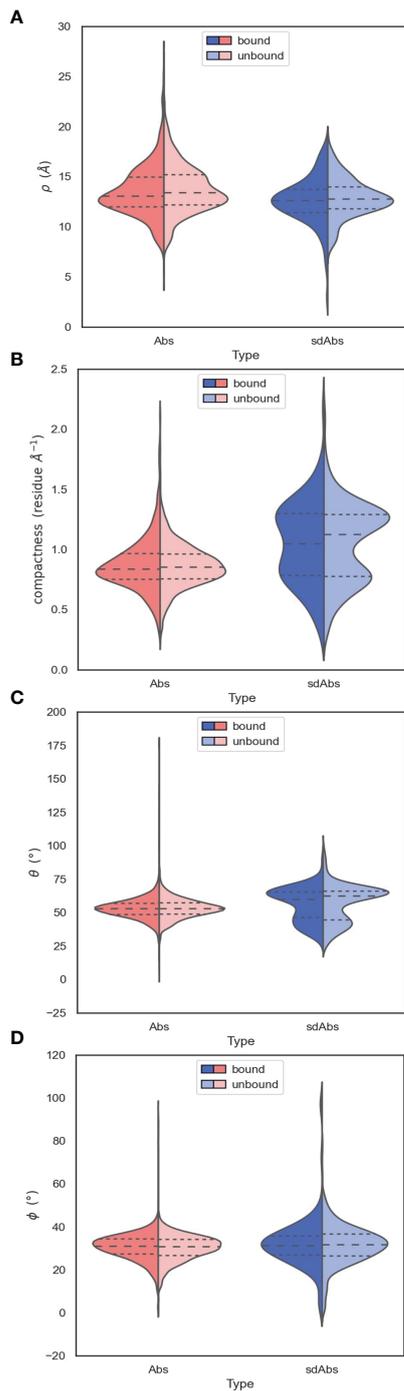


FIGURE 7

The orientation of the CDR-H3 loops of sdAbs suggests why there is overlap in the accessibility of the epitopes targeted by sdAbs and Abs. (A) Distributions of  $\rho$  values show that sdAbs and Abs have similar reach. (B) On average, sdAb CDR-H3 loops are more compacted than Ab loops. (C) Distributions of  $\theta$  values indicate that the majority of sdAb CDR-H3 loops do not extend upwards away from the VHH domain, but lie flat against it. (D) Distributions of  $\phi$  values indicate that the majority of sdAb CDR-H3 loops pack against the VHH domain. In all figures the bound examples are shown in a darker shade on the left of the distributions, with the unbound in a lighter shade on the right.

interactions (Supplementary Figures S15C, S15D). When comparing the paratope of sdAbs only to the paratope residues

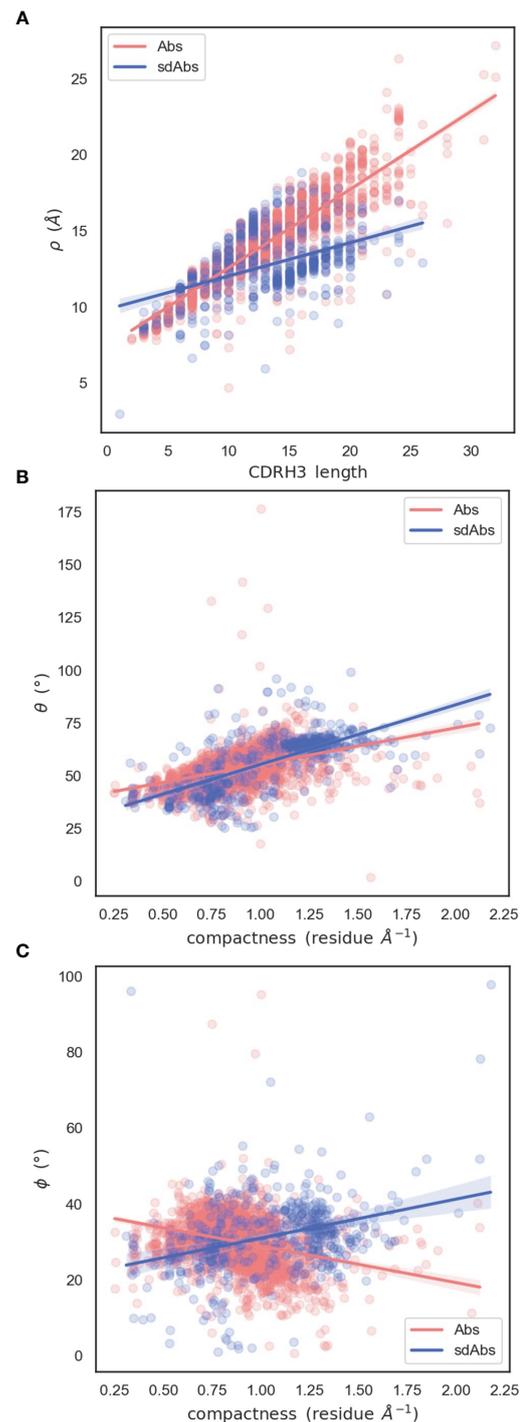


FIGURE 8

Relationships between spherical angles and compactness suggest that the paratope is stabilized by the CDR-H3 loop packing against VL domains in Abs, or the rest of the VHH domain in sdAbs. (A) Correlation between  $\rho$  and CDR-H3 length (B) Correlation between  $\theta$  and compactness (C) Correlation between  $\phi$  and compactness.

from the Ab VH domain, again significant differences are found (Supplementary Figures S15E, S15F). We observe a minimal number of examples where the CDR-H3 loop contributes zero interactions (Figure 11). These results show that the highly variable CDR-H3 loop is even more dominant in sdAbs than in Abs. This,

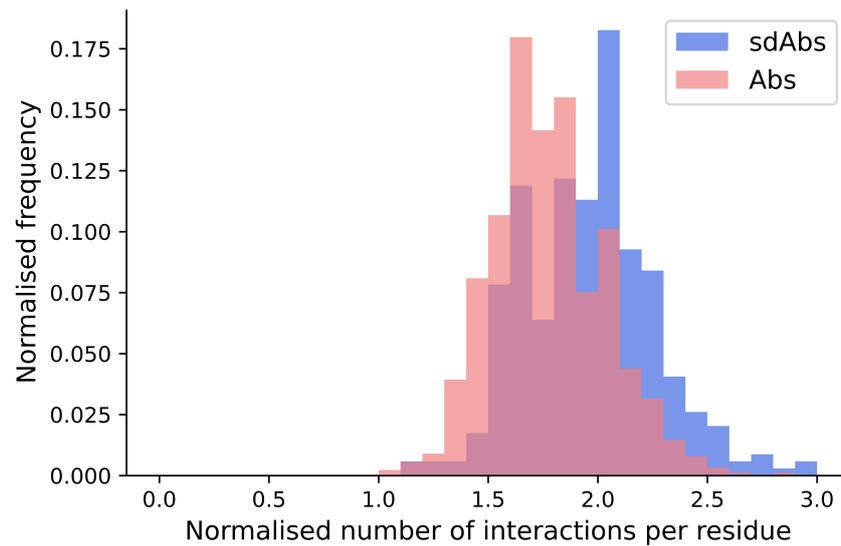


FIGURE 9

The distributions of the number of interactions initiated by sdAbs (blue) and Abs (pink) paratopes demonstrate that sdAb paratopes establish significantly more interactions per residue than Ab paratopes. Comparing the number of interactions from sdAbs to Abs, normalized for paratope size, we find a mean increase of 0.19.

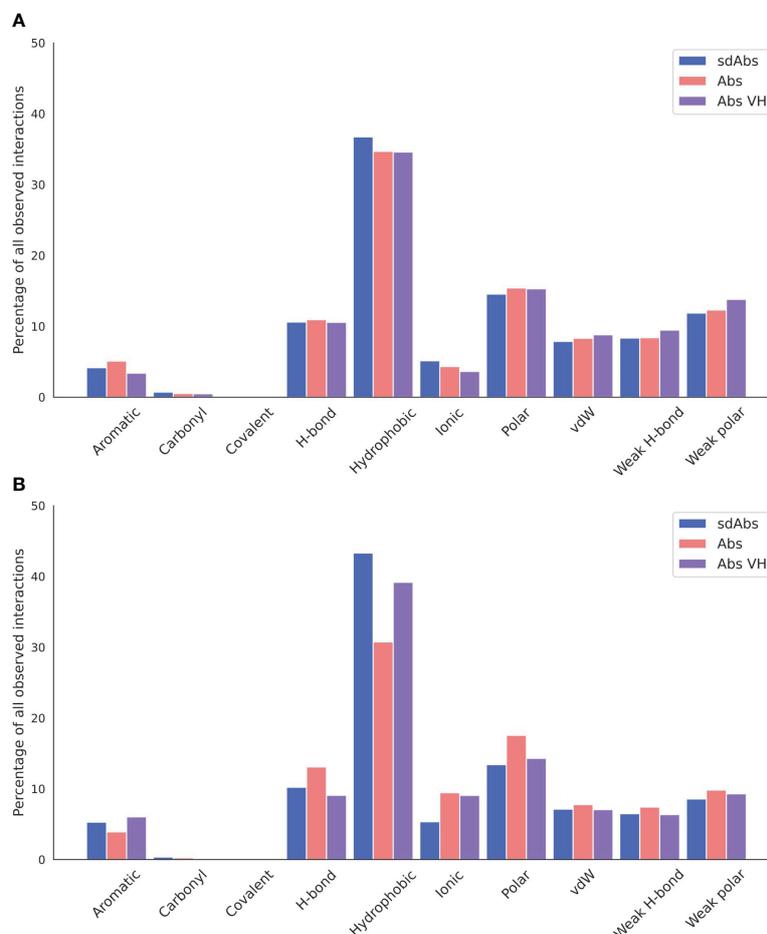


FIGURE 10

Hydrophobic interactions dominate across sdAb-Ag and Ab-Ag complexes. Total occurrences as a percentage of all interaction types observed for the (A) CDR loops and (B) the framework regions. Results for sdAbs are shown in blue, Abs are shown in pink and the VH domain of Abs are shown in purple.

however, is not the only difference: we also observe that the paratopes of sdAbs tend to contain a smaller proportion of CDR residues than Abs (Figure 12, Supplementary Figure S16), from which we can infer that sdAbs show greater inclusion of framework residues in their paratopes than Abs.

### 3.15 Interacting framework residues are often conserved in sdAbs

Given we find that framework residues make up a larger proportion of the paratope in sdAbs than in Abs (Figure 12), we next tested if these framework residues show high variability, undergoing somatic hypermutation to improve binding, or are conserved germline residues.

Framework residues observed in the interactions-defined paratope in at least 10% of the sdAb complexes were determined (Supplementary Table S3) and in descending order of frequency, include positions 66 (50.4%), 52 (31.6%), 55 (27.2%), 42 (24.1%), 50 (17.4%), 118 (15.9%), 69 (12.8%), 67 (12.8%), 40 (10.4%), and 2 (10.1%).

The amino acid compositions of these identified framework positions were determined for both of the structural datasets and for the sequence datasets (Figure 13). Positions were not included if less than 5% of the structures or sequences had a residue at that position. We compare the positions found in the interactions-defined paratopes from the structural datasets to a background composition taken from the sequence datasets. The sequence logo plots (Figure 13), show similarities between the paratope composition and background particularly for positions 2, 50, 67, 69 and 118 in sdAbs. The low level of variation at these positions in

sdAbs indicates they are conserved and suggests that they may not contribute to binding specificity.

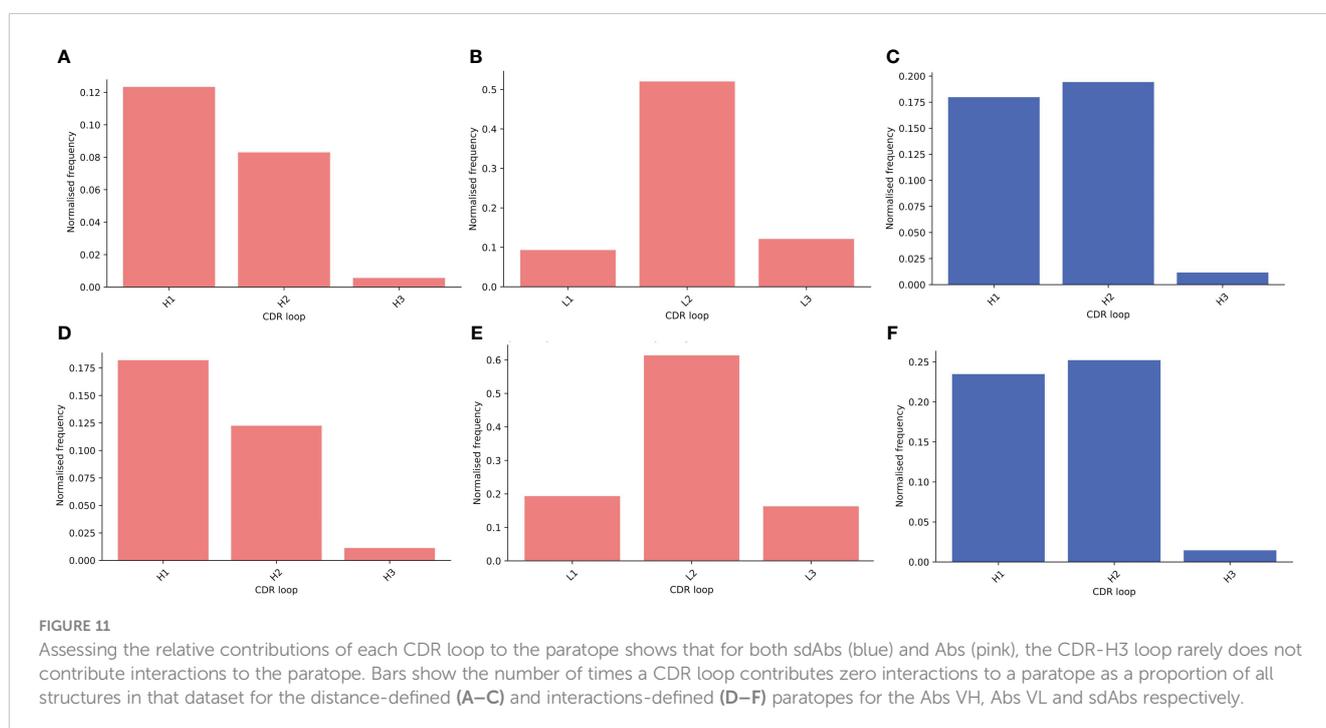
### 3.16 Abs and sdAbs can bind the same epitopes but interact with them differently

Our results suggest that Abs and sdAbs can engage similar types of epitopes but use different mechanisms to do so. Here, we compare the features of an Ab (PDB ID: 6YLA) and a sdAb (PDB ID: 6WAQ) that both bind to the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein, using interactions-defined binding sites.

The sdAb has a longer CDR-H3 (18 residues) than the Ab (12 residues) and the sdAb paratope is smaller than that of the Ab (15 compared to 26 residues). The sdAb paratope includes framework positions 66 and 69, both of which we found to be commonly part of sdAb paratopes. The Ab paratope includes framework positions 1 from the heavy chain and position 68 from the light chain.

Despite the differences in the sdAb and Ab paratopes, they are binding a very similar epitope (Figure 14). The epitopes on the RBD that these structures bind are of a similar size (15 residues for the Ab epitope and 18 residues for the sdAb epitope).

Thirty-one total interactions occur between the Ab epitope and paratope, whilst there are twenty-nine for the sdAb binding site, however when we consider the size of the paratope, this results in an average of 1.9 interactions per paratope residue for the sdAb, compared to 1.2 per Ab paratope residue. In addition, the CDR-H3 has increased importance for the sdAb binding activity. For the Ab, 6 out of the 26 residues in the paratope come from the CDR-H3 loop, whereas for the sdAb, it is 9 out of 15.



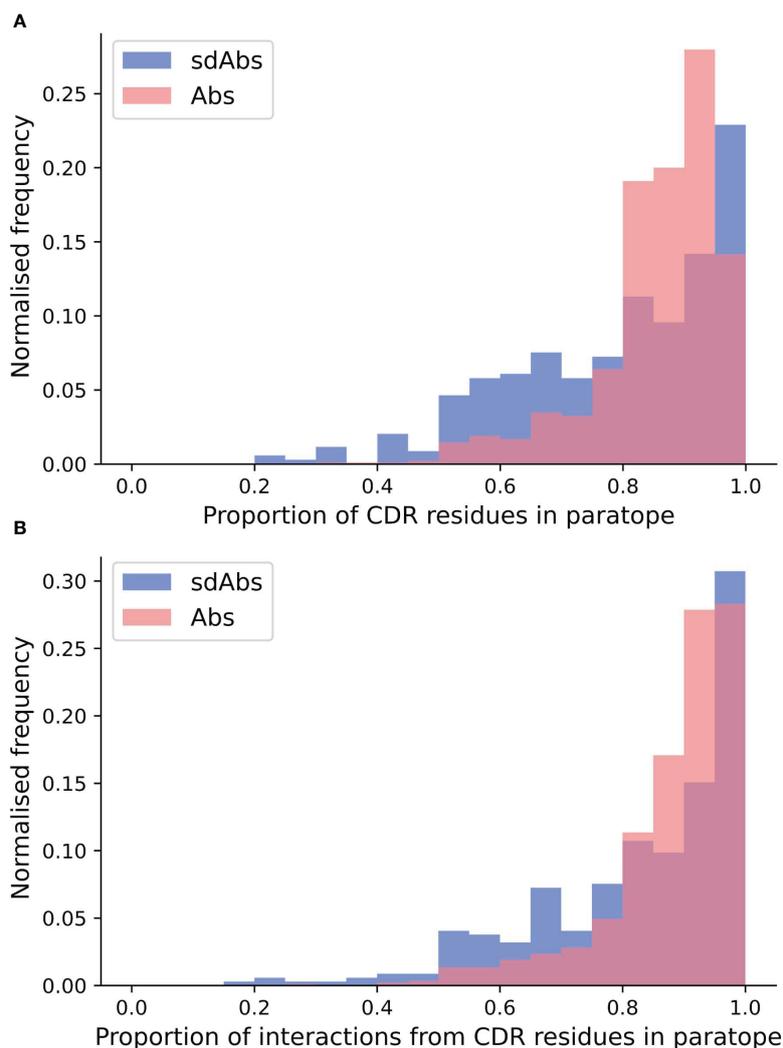


FIGURE 12

Distributions of (A) the proportion of CDR residues in the paratope and (B) the proportion of interactions from CDR residues across the whole paratope, determined per complex in the sdAbs (blue) and Abs (pink) datasets. Higher density on the lower end for the sdAb dataset (blue), compared to the Ab dataset (pink), indicates that more framework residues are involved in binding the epitope.

## 4 Discussion

In this study, we compared the binding sites of sdAbs and Abs to assess whether these two modalities may be suited to different types of epitopes. Overall we find that the paratopes of sdAbs and Abs have distinguishable characteristics. Paratopes of sdAbs tend to be smaller, the CDR conformations observed are different between sdAbs and Abs, and sdAbs tend to have longer CDR-H3 loops than their Ab counterparts. These results are all consistent with previous studies on smaller datasets (6, 12, 17).

These differences in their paratopes led to the expectation that Abs and sdAbs would bind distinct types of epitopes. However, we find that, apart from the epitopes of Abs being slightly more linear than those of sdAbs, the epitopes targeted by sdAbs and Abs cannot be easily distinguished. SdAbs and Abs target epitopes of similar size, similar amino acid compositions and similar accessibility.

There are several suggestions in the literature that the longer CDR-H3 loop of a sdAb means it can interact with epitopes that are

less accessible to conventional Abs by protruding into the cavity (13–15). Henry and MacKenzie (2018) (16) stress that despite individual case studies supporting this hypothesis, the evidence that sdAbs preferentially bind more cryptic epitopes is limited and it is unknown whether this is a general trend across sdAbs. We find that overall, for our datasets, the epitopes targeted by sdAbs are slightly (but significantly) less accessible than epitopes targeted by Abs. However, the absolute difference is small. Furthermore, we find no correlation between CDR-H3 loop length and epitope accessibility.

These results are supported by our finding that Ab and sdAb CDR-H3 loops show differences in their orientation relative to the rest of the supporting VH/VL or VHH domain. We find that sdAb CDR-H3 loops are more compacted than Ab loops and are often found packed against the rest of the VHH domain. For Abs, orientation of the CDR-H3 away from the VH domain leads to its positioning towards the VL domain. As the presence of the VL domain provides steric hindrance, the CDR-H3 loop is forced into a

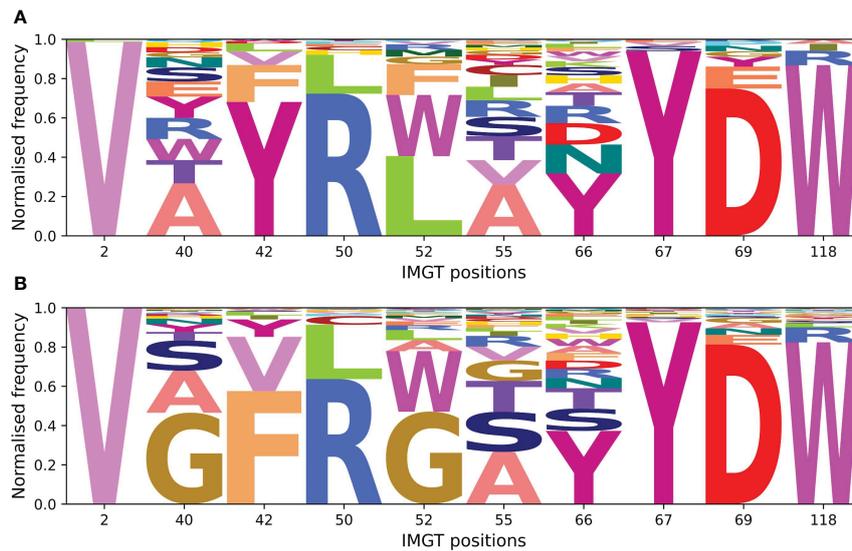


FIGURE 13

Sequence logo plots for framework positions often involved in the paratopes of Abs and sdAbs suggest that framework residues identified to often occur in the paratope are highly conserved in sdAbs. (A) Amino acid compositions at positions found in at least 10% of sdAbs paratopes in our sdAbs structural dataset. (B) Background amino acid compositions in our sdAbs sequence dataset for positions found in at least 10% of sdAbs paratopes. Positions were not included if less than 5% of sequences had a residue at the given position.

conformation that orients it away from the Ab, therefore reducing compactness and increasing reach. In contrast, for sdAbs, orientation of the CDR-H3 away from the VH domain leads to positioning towards empty space and therefore packing against the rest of the VHH domain. These results offer a possible explanation

for our observation that the longer CDR-H3 loops of sdAbs do not necessarily target deeper epitopes.

In addition, we observe that framework residues are more often observed in the paratopes of sdAbs. The importance of framework residues in sdAbs has been indicated in several studies (6, 12, 35, 36).

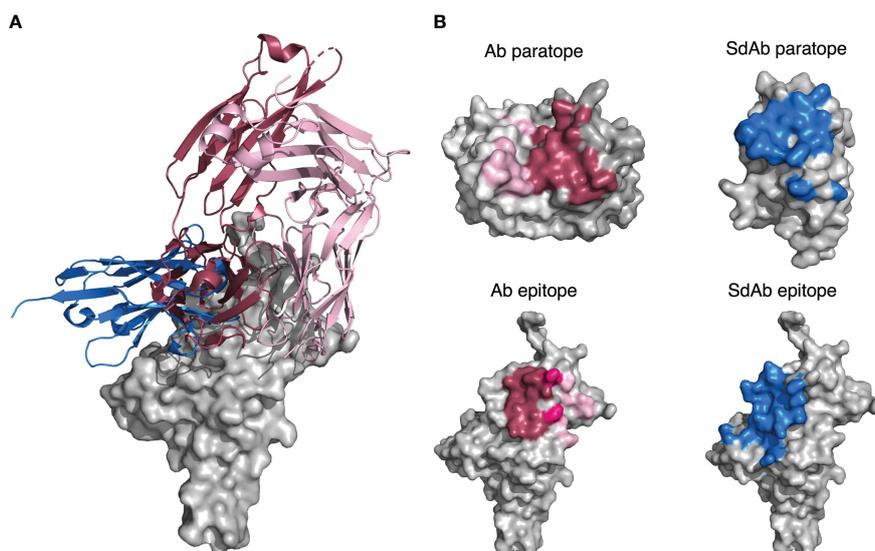


FIGURE 14

(A) A sdAb (PDB ID: 6WAQ) and Ab (PDB ID: 6YLA) are able to bind the SARS-CoV-2 RBD with overlapping epitopes. Dark pink cartoon = Ab heavy chain, light pink cartoon = Ab light chain, blue cartoon = sdAb, grey = surface representation of the SARS-CoV-2 RBD. (B) Abs in general have larger paratopes than sdAbs, but sdAbs are able to bind similarly-sized epitopes as exemplified by structures 6YLA (Ab) and 6WAQ (sdAb). The surface of the Ab heavy chain is shown in dark grey and the light chain in light grey, where the dark pink region represents paratope residues contributed by the VH and the light pink region represents paratope residues contributed by the VL. The surface of the sdAb is shown in light grey with the blue region representing the sdAb paratope residues. The surface of the SARS-CoV-2 antigen is shown in light grey for both the sdAb and Ab, where the Ab epitope is colored dark pink where it is targeted by the Ab VH, light pink where it is targeted by the Ab VL, and a medium pink where it is targeted by both chains. The sdAb epitope is shown in blue. The antigen structures from each PDB were merged to create a complete image of the antigen for the sdAb.

This increase in framework residues is likely related to their increased accessibility due to the lack of the VL domain. Indeed, our results show that most of the framework positions observed in more than 10% of the sdAbs paratopes are frequently observed in the VH-VL interface of Abs (37). Most of the framework positions commonly involved in binding in sdAbs belong to FR2, which is identified by both Zavrtnik et al. (2018) (6) and Mitchell and Colwell (2018a) (12) as an important region for antigen binding. The majority of our identified potential paratope framework residues appear to be highly conserved. Our findings that sdAb CDR-H3 loops often pack against the VHH domain, and that FR2 residues are often conserved, is in agreement with that of Sang et al. (2022) (36), who find that the longer CDR-H3 loops of sdAbs can fold back to interact with FR2 residues.

Finally, we also find that despite tending to have smaller paratopes, sdAbs are able to target similarly-sized epitopes to Abs. This may be explained by our finding that the CDR-H3 loops of sdAbs make a significantly greater number of interactions with the epitope per loop residue than those of Abs, even after normalizing by loop length. Given that these may include conserved framework residues, that will contribute to binding affinity but not specificity, this raises important questions over the specificity of the sdAb binding site, as well as having implications for engineering therapeutics.

## 5 Conclusions

Overall, this study highlights structural characteristics of sdAbs pertinent to the design and engineering of sdAb therapeutics, and calls attention to the need for additional criteria when deciding on the best modality for a particular epitope.

## Data availability statement

The code generated and datasets analyzed for this study can be found at [github.com/oxpig](https://github.com/oxpig). Further inquiries can be directed to the corresponding author.

## Author contributions

CD conceptualized and designed the study. ER, RS, and CD supervised the project. GG, HC, and BG carried out the data curation and analysis. The manuscript was written by GG, with contributions from BG and HC, and reviewed by ER, RS, and CD. All authors contributed to the article and approved the submitted version.

## References

- Chames P, Regenmortel MV, Weiss E, Baty D. Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol* (2009) 157(2):220–33. doi: 10.1111/j.1476-5381.2009.00190.x
- Krah S, Schröter C, Zielonka S, Empting M, Valldorf B, Kolmar H. Single-domain antibodies for biomedical applications. *Immunopharmacol Immunotoxicology* (2015) 38(1):21–8. doi: 10.3109/08923973.2015.1102934
- Yang EY, Shah K. Nanobodies: next generation of cancer diagnostics and therapeutics. *Front Oncol* (2020) 10:1182. doi: 10.3389/fonc.2020.01182
- Jovčevska I, Muyldermans S. The therapeutic potential of nanobodies. *BioDrugs* (2019) 34(1):11–26. doi: 10.1007/s40259-019-00392-z
- Muyldermans S. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* (2013) 82:775–97. doi: 10.1146/annurev-biochem-063011-092449
- Zavrtnik U, Lukan J, Loris R, Lah J, Hadži S. Structural basis of epitope recognition by heavy-chain camelid antibodies. *J Mol Biol* (2018) 430(1):4369–86. doi: 10.1016/j.jmb.2018.09.002
- Bannas P, Hambach J, Koch-Nolte F. Nanobodies and nanobody-based human heavy chain antibodies as antitumor therapeutics. *Front Immunol* (2017) 8:1603. doi: 10.3389/fimmu.2017.01603
- Czajka TF, Vance DJ, Mantis NJ. Slaying SARS-CoV-2 one (single-domain) antibody at a time. *Trends Microbiol* (2021) 29(3):195–203. doi: 10.1016/j.tim.2020.12.006

## Funding

This work was supported by the Engineering and Physical Sciences Research Council (grant number EP/S024093/1), the Wellcome Trust (grant number 102164/Z/13/Z) and the Medical Research Council (grant number MR/R015708/1). The authors also declare that this study received funding from Twist Bioscience. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

## Acknowledgments

The authors thank Claire Marks and Katya Putintseva from LabGenius for their work on the project.

## Conflict of interest

Author RS was employed by company Twist Bioscience.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1231623/full#supplementary-material>

9. Vincke C, Loris R, Saerens D, Martinez-Rodriguez S, Muyldermans S, Conrath K. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J Biol Chem* (2009) 284(4):3273–84. doi: 10.1074/jbc.m806889200
10. Muyldermans S, Baral T, Retamozzo VC, Baetselier PD, Genst ED, Kinne J, et al. Camelid immunoglobulins and nanobody technology. *Veterinary Immunol Immunopathology* (2009) 128(1–3):178–83. doi: 10.1016/j.vetimm.2008.10.299
11. Sircar A, Sanni KA, Shi J, Gray JJ. Analysis and modeling of the variable region of camelid single-domain antibodies. *J Immunol* (2011) 186(11):6357–67. doi: 10.4049/jimmunol.1100116
12. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. *Proteins: Structure Function Bioinf* (2018) 86(7):697–706. doi: 10.1002/prot.25497
13. Genst ED, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci* (2006) 103(12):4586–91. doi: 10.1073/pnas.0505379103
14. Wesolowski J, Alzogaray V, Reyelt J, Unger M, Juarez K, Urrutia M, et al. Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. *Med Microbiol Immunol* (2009) 198:157–74. doi: 10.1007/s00430-009-0116-7
15. Desmyter A, Spinelli S, Roussel A, Cambillau C. Camelid nanobodies: killing two birds with one stone. *Curr Opin Struct Biol* (2015) 32:1–8. doi: 10.1016/j.sbi.2015.01.001
16. Henry KA, MacKenzie CR. Antigen recognition by single-domain antibodies: structural latitudes and constraints. *mAbs* (2018) 10(6):815–26. doi: 10.1080/19420862.2018.1489633
17. Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Engineering Design Selection* (2018) 31(7–8):267–75. doi: 10.1093/protein/gzy017
18. Schneider C, Raybould MIJ, Deane CM. SABDab in the age of biotherapeutics: updates including SABDab-nano, the nanobody structure tracker. *Nucleic Acids Res* (2021) 50(D1):D1368–72. doi: 10.1093/nar/gkab1050
19. Eliyahu S, Sharabi O, Elmedvi S, Timor R, Davidovich A, Vigneault F, et al. Antibody repertoire analysis of hepatitis c virus infections identifies immune signatures associated with spontaneous clearance. *Front Immunol* (2018) 9. doi: 10.3389/fimmu.2018.03004
20. Li X, Duan X, Yang K, Zhang W, Zhang C, Fu L, et al. Comparative analysis of immune repertoires between bactrian camel's conventional and heavy-chain antibodies. *PLoS One* (2016) 11(9):e0161801. doi: 10.1371/journal.pone.0161801
21. Olsen TH, Boyles F, Deane CM. Observed antibody space: a diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci* (2021) 31(1):141–6. doi: 10.1002/pro.4205
22. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SABDab: the structural antibody database. *Nucleic Acids Res* (2013) 42(D1):D1140–6. doi: 10.1093/nar/gkt1043
23. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* (2006) 22(13):1658–9. doi: 10.1093/bioinformatics/btl158
24. Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and t cell receptor variable domains and ig superfamily v-like domains. *Dev Comp Immunol* (2003) 27(1):55–77. doi: 10.1016/s0145-305x(02)00039-3
25. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* (2015) 32(2):btv552. doi: 10.1093/bioinformatics/btv552
26. Jubb HC, Higuero AP, Ochoa-Montano B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J Mol Biol* (2017) 429(3):365–71. doi: 10.1016/j.jmb.2016.12.004
27. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* (2009) 25(11):1422–3. doi: 10.1093/bioinformatics/btp163
28. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 2.4.1, Schrödinger, LLC. (2015). Available at: <https://pymol.org/2/support.html?#citing>
29. Kelow S, Faezov B, Xu Q, Parker M, Adolf-Bryfogle J, Roland L Dunbrack J. A penultimate classification of canonical antibody cdr conformations. *bioRxiv* (2022). doi: 10.1101/2022.10.12.511988
30. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. scpUCSF ChimeraX/scp: structure visualization for researchers, educators, and developers. *Protein Sci* (2020) 30(1):70–82. doi: 10.1002/pro.3943
31. Hunter JD. Matplotlib: a 2d graphics environment. *Computing Sci & Eng* (2007) 9(3):90–5. doi: 10.1109/mcse.2007.55
32. Wong WK, Robinson SA, Bujotzek A, Georges G, Lewis AP, Shi J, et al. Ability: identifying sequence-dissimilar antibodies that bind to the same epitope. *mAbs* (2021) 13(1):1873478. doi: 10.1080/19420862.2021.1873478
33. Sela M, Schechter B, Schechter I, Borek F. Antibodies to sequential and conformational determinants. *Cold Spring Harbor Symp Quantitative Biol* (1967) 32:537–45. doi: 10.1101/sqb.1967.032.01.065
34. Forsström B, Axnäs BB, Rockberg J, Danielsson H, Bohlin A, Uhlen M. Dissecting antibodies with regards to linear and conformational epitopes. *PLoS One* (2015) 10(3):e0121673. doi: 10.1371/journal.pone.0121673
35. Kelow SP, Adolf-Bryfogle J, Dunbrack RL. Hiding in plain sight: structure and sequence analysis reveals the importance of the antibody DE loop for antibody-antigen binding. *mAbs* (2020) 12(1):1840005. doi: 10.1080/19420862.2020.1840005
36. Sang Z, Xiang Y, Bahar I, Shi Y. Llamade: an open-source computational pipeline for robust nanobody humanization. *Structure* (2022) 30(3):418–429.e3. doi: 10.1016/j.str.2021.11.006
37. Raybould MIJ, Marks C, Kovaltsuk A, Lewis AP, Shi J, Deane CM. Public baseline and shared response structures support the theory of antibody repertoire functional commonality. *PLoS Comput Biol* (2021) 17(3):e1008781. doi: 10.1371/journal.pcbi.1008781